wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 02:15 PM EDT

PDB ID	:	3CKD
Title	:	Crystal structure of the C-terminal domain of the Shigella type III effector IpaH
Authors	:	Lam, R.; Singer, A.U.; Cuff, M.E.; Skarina, T.; Kagan, O.; DiLeo, R.; Edwards, A.M.; Joachimiak, A.; Savchenko, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2008-03-14
Resolution	:	2.65 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

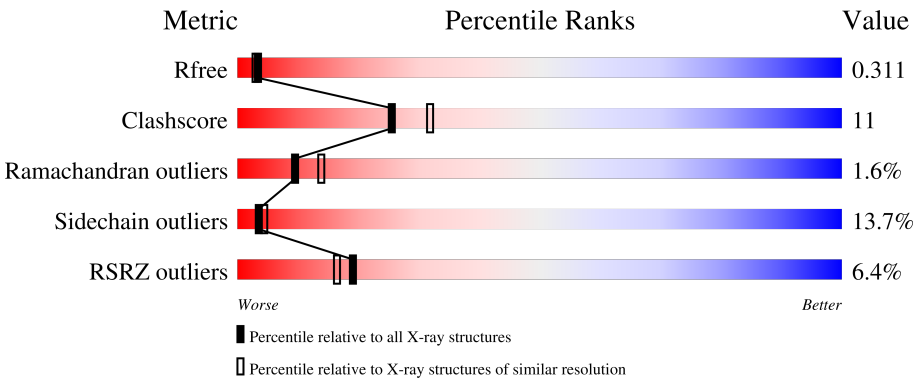
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.65 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	312	7% 55% 23% 6% 16%
1	B	312	8% 61% 12% 5% 22%
1	C	312	% 61% 22% • • 13%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6327 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

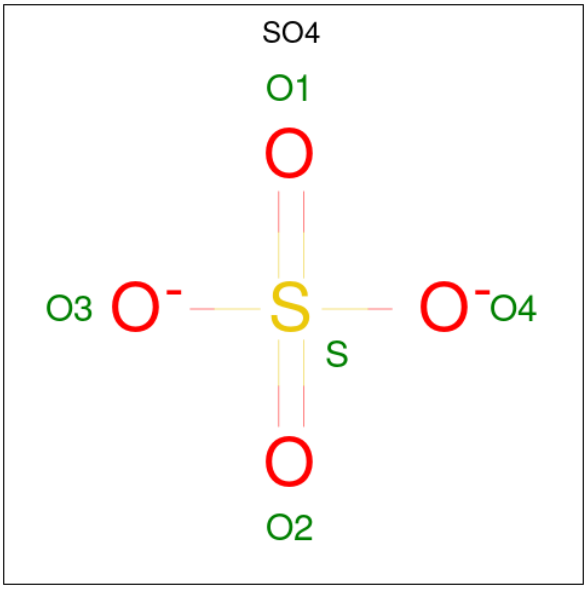
- Molecule 1 is a protein called Invasion plasmid antigen, secreted by the Mxi-Spa secretion machinery.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	262	2114	1319	373	416	1	5	0	0	0
			Total	C	N	O		Se			
1	B	244	1970	1233	343	389		5	0	0	0
			Total	C	N	O	S	Se			
1	C	272	2184	1359	385	434	1	5	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

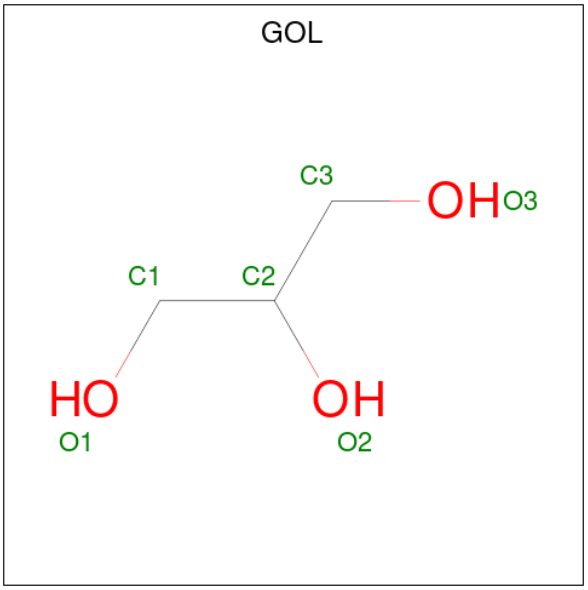
Chain	Residue	Modelled	Actual	Comment	Reference
A	264	GLY	-	EXPRESSION TAG	UNP Q8VSA1
B	264	GLY	-	EXPRESSION TAG	UNP Q8VSA1
C	264	GLY	-	EXPRESSION TAG	UNP Q8VSA1

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



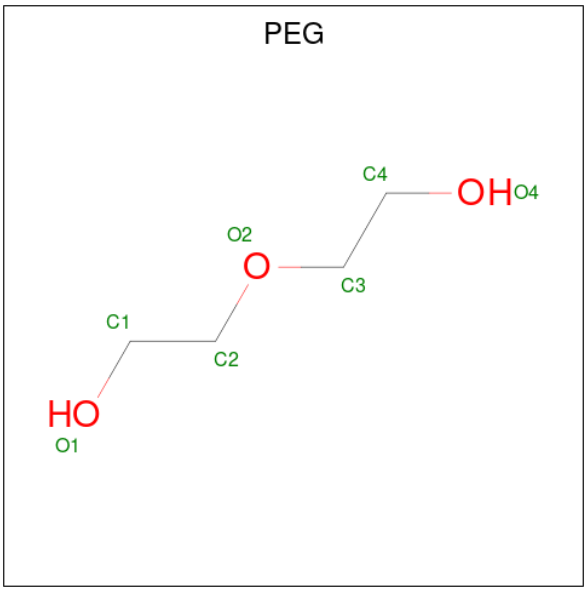
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	B	1	5	4	1	0	0
			Total	O	S		
2	C	1	5	4	1	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	C	1	6	3	3	0	0
			Total	C	O		
3	C	1	6	3	3	0	0
			Total	C	O		
3	C	1	6	3	3	0	0

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	C	1	7	4	3	0	0

- Molecule 5 is water.

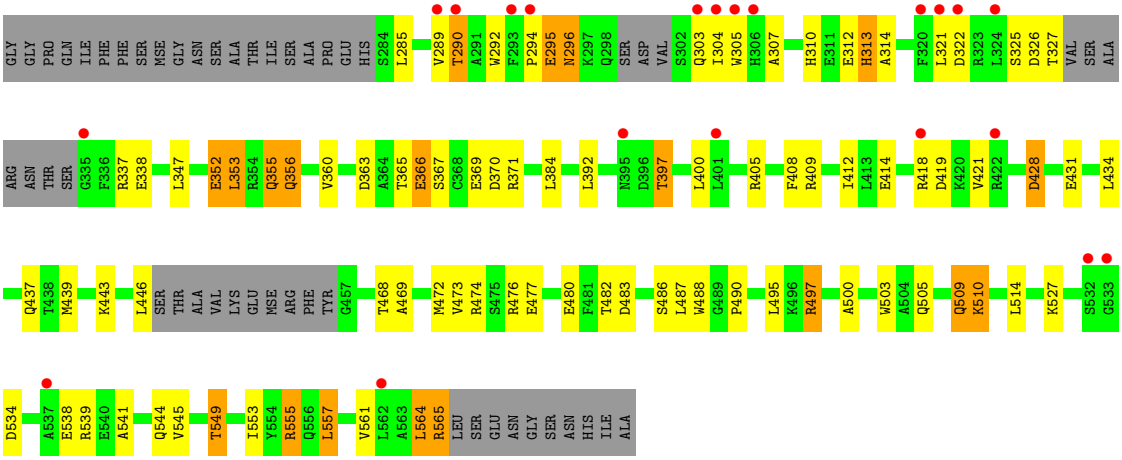
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	1	1	1	0	0
			Total	O		
5	B	4	4	4	0	0
			Total	O		
5	C	19	19	19	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

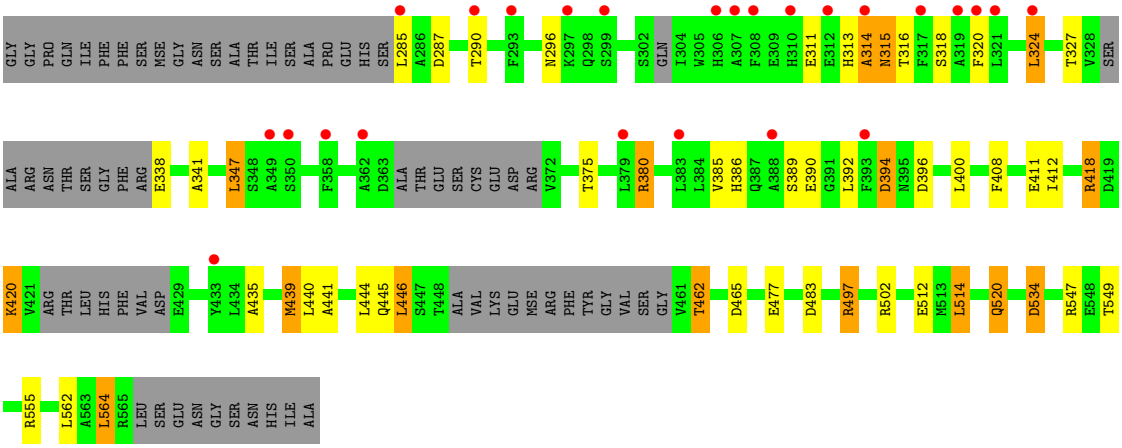
• Molecule 1: Invasion plasmid antigen, secreted by the Mxi-Spa secretion machinery

Chain A:



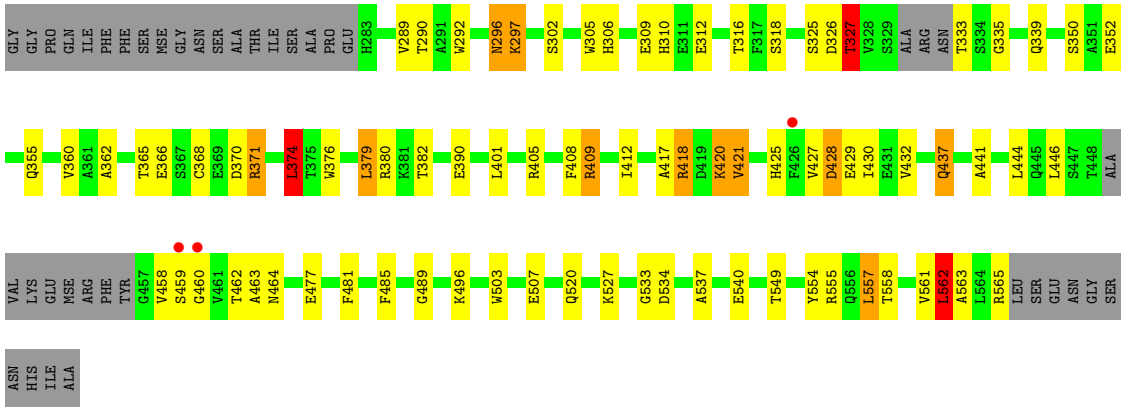
• Molecule 1: Invasion plasmid antigen, secreted by the Mxi-Spa secretion machinery

Chain B:



• Molecule 1: Invasion plasmid antigen, secreted by the Mxi-Spa secretion machinery

Chain C:


GLY GLY PRO GLN ILE PHE PHE SER MSE GLY ASN SER ALA THR ILE SER ALA PRO GLU H283 V289 T290 A291 W292 N296 K297 S302 W305 H306 E309 H310 E311 E312 T316 F317 S318 S325 D326 T327 V328 S329 ALA ARG ASN T333 S334 G335 Q339 S350 A351 E352
Q355 V360 A361 A362 T365 E366 S367 C368 E369 D370 R371 L374 T375 W376 L379 R380 K381 T382 E390 L401 R405 F408 R409 I412 A417 R418 D419 K420 V421 H425 F426 V427 D428 E429 I430 E431 V432 Q437 A441 L444 Q445 L446 S447 T448 ALA
VAL LYS GLU MSE ARG PHE TYR G457 V458 S459 G460 V461 T462 A463 N464 E477 F481 F485 G489 K496 W503 E507 Q520 K527 G533 D534 A537 E540 T549 Y554 R555 Q556 L557 T558 V561 L562 A563 L564 R565 LEU SER GLU ASN GLY SER
ASN HIS ILE ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α, β, γ	128.93Å 128.93Å 282.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.21 – 2.65 49.19 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.0 (49.21-2.65) 98.9 (49.19-2.65)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.49 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.228 , 0.284 0.261 , 0.311	Depositor DCC
R_{free} test set	1743 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	58.5	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 49.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	6327	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PEG, SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	A	0.65	0/2145	0.72	1/2889 (0.0%)
1	B	0.71	0/1996	0.75	3/2688 (0.1%)
1	C	1.04	7/2217 (0.3%)	0.97	8/2990 (0.3%)
All	All	0.83	7/6358 (0.1%)	0.82	12/8567 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	481	PHE	CE1-CZ	5.96	1.48	1.37
1	C	507	GLU	CG-CD	5.52	1.60	1.51
1	C	485	PHE	CE1-CZ	5.36	1.47	1.37
1	C	390	GLU	CB-CG	-5.33	1.42	1.52
1	C	540	GLU	CG-CD	5.30	1.59	1.51

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	409	ARG	NE-CZ-NH2	-8.23	116.19	120.30
1	C	565	ARG	NE-CZ-NH2	-7.70	116.45	120.30
1	C	409	ARG	NE-CZ-NH1	7.14	123.87	120.30
1	C	418	ARG	NE-CZ-NH1	6.39	123.49	120.30
1	C	374	LEU	CA-CB-CG	6.03	129.17	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	368	CYS	Peptide
1	C	533	GLY	Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2114	0	2017	64	0
1	B	1970	0	1885	26	0
1	C	2184	0	2081	41	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
3	C	18	0	24	1	0
4	C	7	0	10	1	0
5	A	1	0	0	0	0
5	B	4	0	0	0	0
5	C	19	0	0	0	0
All	All	6327	0	6017	130	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 11.

The worst 5 of 130 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:555:ARG:HH11	1:A:555:ARG:CG	1.56	1.16
1:A:555:ARG:HH11	1:A:555:ARG:HG3	1.08	1.11
1:A:303:GLN:HE21	1:A:304:ILE:HG13	1.19	1.06
1:A:555:ARG:HG3	1:A:555:ARG:NH1	1.71	0.97
1:A:468:THR:HG22	1:A:472:MSE:HE2	1.54	0.89

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	254/312 (81%)	219 (86%)	31 (12%)	4 (2%)	9 14
1	B	232/312 (74%)	207 (89%)	20 (9%)	5 (2%)	6 9
1	C	266/312 (85%)	247 (93%)	16 (6%)	3 (1%)	14 21
All	All	752/936 (80%)	673 (90%)	67 (9%)	12 (2%)	9 14

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	313	HIS
1	B	394	ASP
1	A	370	ASP
1	B	314	ALA
1	B	564	LEU

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	218/252 (86%)	183 (84%)	35 (16%)	2 2
1	B	204/252 (81%)	177 (87%)	27 (13%)	4 5
1	C	228/252 (90%)	201 (88%)	27 (12%)	5 7
All	All	650/756 (86%)	561 (86%)	89 (14%)	3 4

5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	520	GLN
1	C	352	GLU
1	B	547	ARG
1	C	309	GLU
1	C	371	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	395	ASN
1	B	437	GLN
1	C	437	GLN
1	C	283	HIS
1	C	296	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

6 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	C	600	-	4,4,4	0.23	0	6,6,6	0.58	0
3	GOL	C	603	-	5,5,5	0.45	0	5,5,5	0.43	0
3	GOL	C	604	-	5,5,5	0.26	0	5,5,5	0.86	0
4	PEG	C	605	-	6,6,6	0.66	0	5,5,5	0.56	0
3	GOL	C	602	-	5,5,5	0.42	0	5,5,5	0.63	0
2	SO4	B	601	-	4,4,4	0.22	0	6,6,6	0.29	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	C	603	-	-	0/4/4/4	-
3	GOL	C	602	-	-	2/4/4/4	-
3	GOL	C	604	-	-	4/4/4/4	-
4	PEG	C	605	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	602	GOL	C1-C2-C3-O3
3	C	604	GOL	O1-C1-C2-C3
3	C	604	GOL	C1-C2-C3-O3
4	C	605	PEG	O1-C1-C2-O2
3	C	602	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	603	GOL	1	0
4	C	605	PEG	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	257/312 (82%)	0.74	21 (8%)	11	9	63, 81, 105, 113	22 (8%)
1	B	239/312 (76%)	0.69	25 (10%)	6	4	46, 62, 81, 89	9 (3%)
1	C	267/312 (85%)	0.16	3 (1%)	80	79	24, 36, 67, 79	9 (3%)
All	All	763/936 (81%)	0.52	49 (6%)	19	16	24, 63, 93, 113	40 (5%)

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	395	ASN	4.9
1	A	305	TRP	4.7
1	A	293	PHE	4.6
1	B	350	SER	4.1
1	B	285	LEU	4.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	PEG	C	605	7/7	0.87	0.21	43,52,61,64	0
3	GOL	C	603	6/6	0.89	0.22	49,53,54,55	6
3	GOL	C	604	6/6	0.90	0.18	63,65,67,72	0
3	GOL	C	602	6/6	0.91	0.26	44,46,46,47	6
2	SO4	B	601	5/5	0.96	0.16	55,56,58,59	5
2	SO4	C	600	5/5	0.98	0.11	48,48,51,51	5

6.5 Other polymers (i)

There are no such residues in this entry.